DOCUMENT RESUME

ED 260 493 EA 017 920

AUTHOR Hansen, Kenneth H.
TITLE State Evaluation and Assessment Programs: SEA Policy Options. Discussion Draft.
INSTITUTION Northwest Regional Educational Lab., Portland, OR. Northwest Center for State Educational Policy Studies.
SPONS AGENCY Chief State School Officers of the Northwest and Pacific.; National Inst. of Education (ED), Washington, DC.
PUB DATE Mar 85
CONTRACT 400-83-0005
NOTE 16p.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Agency Cooperation; Data Interpretation; *Educational Assessment; Educational Policy; Measurement Objectives; Measurement Techniques; *Policy Formation; Selection; State Departments of Education; *State Programs

ABSTRACT

When establishing statewide and nationally comparable educational evaluation and assessment systems, state education agencies (SEAs) must consider factors affecting policy in at least four areas. First, the full range of purposes for establishing such a system should be clarified. It must be decided whether the system exists to provide data for reporting or for decision-making, whether comparisons should be made within educational units or between them, and what emphasis is to be placed on various forms of learning. Second, the evaluation tools selected must be compatible with the purposes identified. The range of evaluation information gathered and the contextual data collected must be appropriate, and attention should be paid to whether the tools used are suitably diagnostic and reveal significant facts. The third area of concern is the potential for misusing or misinterpreting the data. Full disclosure of a broad range of data, coupled with thorough explanation of how to understand it, is vital. Fourth, careful consideration must be given to the degree of collaboration that is desired with other agencies, and the extent to which other decisions affect the achievement of the desired cooperation. (PGD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED260493

# STATE EVALUATION AND ASSESSMENT PROGRAMS: SEA POLICY OPTIONS

**Discussion Draft**

Prepared for the
Chief State School Officers of the Northwest and Pacific

Kenneth H. Hansen
Principal Author

Northwest Center for State Educational Policy Studies
Northwest Regional Educational Laboratory
300 Southwest Sixth Avenue
Portland, Oregon 97204

March 1985

EA 017 920

The work upon which this publication is based was performed pursuant to Contract 400-83-0005 of the National Institute of Education. It does not, however, necessarily reflect the views of that agency.

TABLE OF CONTENTS


| | Page |
|---|---|
| INTRODUCTION | 1 |
| I. ESTABLISHING PURPOSES | 2 |
| II. CHOOSING INDICATORS/MEASURES | 5 |
| III. GUARDING AGAINST MISUSE/MISINTERPRETATION | 8 |
| IV. BUILDING COLLABORATIVE RELATIONSHIPS | 9 |
| IN CONCLUSION | 12 |

# STATE EVALUATION AND ASSESSMENT PROGRAMS: SEA POLICY OPTIONS

## INTRODUCTION

The Chief State School Officers of the Northwest and Pacific have asked the NWREL Center for State Studies to set forth some of the policy considerations which may be involved in the implementation of the CCSSO position paper, EDUCATION EVALUATION AND ASSESSMENT IN THE UNITED STATES. The document here is in response to this request, attempting to extend, affirm, and augment the Council paper.

Analysis of the Council position paper, illuminated by conversations with individual Chiefs, has resulted in the identification of four major broad areas in which significant policy issues arise: (1) establishing purposes; (2) choosing indicators/measures; (3) guarding against misuse/misinterpretation; and (4) building collaborative relationships among all levels of educational governance.

In each of these policy areas, and doubtless in others which have not yet become clearly apparent, it seems obvious that the individual SEA will have to play a very active role; the one option no longer realistically available is the "no-action" option. Compounding the problem facing the SEAs is the lack of really clear-cut choices to be made: there are really very few simple yes-or-no, this-or-that options available, but rather matters of relative emphasis, or "tilt" in one direction or another. Nevertheless, it does seem possible to set forth and to clarify some of the policy issues involved.

## I. ESTABLISHING PURPOSES

Single v. multiple purposes. It is becoming very clear that any limited, single purpose for establishing statewide (and nationally comparable) education evaluation and assessment systems is unlikely to meet the needs of the SEAs or the various constituencies to which they report. At the outset of the CCSSO involvement which led to the adoption of the Council position paper there may have been a very primary concern with the "wall chart" approach to comparative state reporting, but the concern now is infinitely broader than that.

Both the original position paper and the subsequent Council proposal for the establishment of a nationwide evaluation and assessment center list many purposes which would be served by the establishment and/or improvement of statewide programs. The very fundamental purposes would be (1) the establishment of a means for the monitoring of the progress in education reforms and (2) for demonstrating accountability for the results of these reforms. Also included as stated purposes of the proposal to improve statewide evaluation and assessment programs are outcomes such as these: to draw public attention to education; to "exhort, motivate, and reward;" to understand better the consequences of change or action; to aid in implementing policies; to examine cost/benefit relationships; to provide records of expenditures; to assist in determining resources and in making resource allocations; and, of course, to allow for reasonable comparisons between and among states.

Any single purpose for establishing or bolstering a state evaluation and assessment program would seem, therefore, to be such a limited choice

as to constitute a less-than-adequate option. A much sounder policy choice would appear to be that of selecting the multiple purposes which meet specific state needs.

Reporting v. decision-making. Another fundamental policy issue facing the states is that of the relative emphasis to be given to designing the system for producing data for reports to various publics or for educational decision-making. Of course, like most of the other policy choices, this is not really an either-or matter, but one of emphasis. Although the initial inclination may well be to collect data to provide to the public as evidence of professional accountability, surely very strong consideration should be given to greater emphasis on accumulating data which is internally useful for bringing about instructional improvement.

External v. internal comparisons. At issue here is another fundamental question of policy emphasis, rather than outright policy choice. Should the data-generation and data-collection emphasis be on comparisons of achievement and progress within the educational unit (school site, district, or state) or between the units? There certainly is merit in being able to compare one unit with another, but there is much greater merit, it would seem, in emphasizing the rate and degree of progress a school, a district, or a state is making in achieving its own goals, rather than concentrating on determining which unit has the higher scores or has achieved some other purported measure of "betterness."

Rote/recall v. higher-order learnings. In establishing the fundamental purposes of an evaluation and assessment system, policy consideration will need to be given to the relative emphasis to be placed

on the measurement of essentially factual knowledge as opposed to the higher-order conceptual skills. With the current emphasis being placed on the mastery of a core of content knowledge, there will be strong pressure to concentrate on measuring and reporting that kind of learning--a task which is easier to perform than is the one of measuring the higher-order learnings. Both kinds of learning are of course important; the point here is not to suggest the precise balance between the two, but to underscore the importance of there being a choice consciously made as assessment and evaluation policy is formulated.

Solely academic v. entire range of outcomes. Regardless of what comparative educational assessments are being made, and regardless of by whom or between whom the comparisons are being made, the tendency has generally been to concentrate the efforts and the reports on academic learnings only. As a result, relatively less attention has been given to measuring and reporting progress in other areas--the aesthetic and affective areas, for example, or evidences of growth and achievement in vocational knowledge, good citizenship, or development of positive self-concepts. Obviously, these latter areas are difficult to measure and tricky to report, but there is need for policy determinations and policy statements which reflect the degree of importance which the policy-formulators attach to measuring and reporting the entire range of learnings which are considered to be important.

## II. CHOOSING INDICATORS/MEASURES

Once the purposes of the evaluation and assessment programs have been determined and clearly set forth as established policy, a further set of policy determinations becomes necessary: what indicators/measures will be used? Answers to that question will necessarily involve a number of subsidiary policy determinations.

Multiplicity to match variety of purposes and uses. If by policy there has been established a range and variety of distinct but related purposes for the evaluation and assessment program, no simple and limited list of indicators/measures would seem to be sufficient. As the Council documents have pointed out, indicators will be needed to display varied inputs, the variety of educational processes used, and the wide range of outcomes obtained. For example, indicators and measures will be needed which describe adequately what it is that students are learning, what changes are taking place in the student population, what trends are developing in the availability of resources, and what educational effects are being seen as a result of policy changes. If policy stresses the importance attached to a variety of types of learning, consistency would require that there also be policies supporting the use of a wide variety of indicators. The specific indicators or measures to be used is a technical question; the breadth of their scope is a policy question of the highest importance.

Use of background variables. If legitimate and useful comparisons are to be made between and among various educational units (school-to-school; district-to-district; state-to-state), a host

of background variables needs to be included and considered. Fair comparisons cannot be made without considering, for example, socio-economic status and other population and fiscal variables such as numbers of impoverished or handicapped students, resources and expenditures, and curricular offerings. Comparisons will be made, willy-nilly, so there would seem to be need for exercising the policy options which would call for the inclusion of the greatest feasible number of background variables to legitimize and facilitate these comparisons.

Inclusion of diagnostic instruments. If the purpose of a statewide evaluation assessments program--or a nationwide one, for that matter--is simply to report on the status of education, measurement instruments without diagnostic qualities would probably suffice. But if, in addition to the wholly rational objective of reporting, there is added the objective of improving education, diagnosis of learning styles and learning difficulties becomes extremely important. Granting that the diagnostic results obtained are for internal use, rather than for "outside" reporting, the importance of diagnosis to educational improvement would seem to call for assessment and evaluation policies which reflect a commitment to include diagnosis as part of the program.

Distinguishing between ideal and reality. It is quite possible to employ indicators which, taken at face value and in isolation, do not really give an accurate picture of some aspect of an educational program. For example, listing "courses available" in a given program without also reporting the extent to which these courses are actually taken can give, however inadvertently, an erroneous impression of the

program. It would seem that sound policy would require the inclusion only of such indicators as would give a realistic, not an idealistic, picture of the status of education within the school, district, or state.

Inclusion of special programs. The paragraphs above cautioned about indicators which gave perhaps too rosy a picture, but the opposite problem of an insufficiently favorable picture can likewise be an issue. Policies regarding what will be measured or reported are sometimes so stringently limited to academic matters that other items of importance in making a fair assessment are neglected. For example, data on GED programs and anti-dropout programs are of significance in assessing the holding power of a school. As a matter of policy, all such indicators need to be included.

Range of achievement measures. Since student achievement as it is popularly understood remains the mostly widely accepted measure of educational progress, it may be appropriate to suggest that it would be sound policy to include a wide range of these achievement measures in the evaluation and assessment program. Straight grade-levels and college-test scores are useful and informative, but needed also may well be special reports on specific skills, on specific core-content learnings, and reports on writing samples and other "production" samples which further illustrate "what the kids are learning."

## III. GUARDING AGAINST MISUSE/MISINTERPRETATION

It might seem at first glance that the problem of the possible misuse or misinterpretation of evaluation and assessment data is not a policy question at all, but simply a technical matter, or a public information matter, or maybe just a simple "PR" problem. On closer examination, however, the policy implications are apparent. One of the greatest deterrents to the adoption of educational policies which would permit or facilitate comparisons has been the suspicion that the comparisons, at any level from individual school to nationwide state-by-state, would not be "fair."

Now that the states, as a matter of policy, are generally moving toward the acceptance and (under proper conditions) even the encouragement of comparisons, some basic educational policies would seem to need consideration.

Promoting understanding of conditions and limitations. If fuller educational data, more comprehensive and more comprehensible, is going to be made more widely available--with the consequently inevitable "comparisons" which have been so long feared and avoided--there will emerge the need for educational policies which directly speak to the problem of misinterpretation. Public understanding will not just have to be desired; it will have to be promoted. Staff time and fiscal resources will have to be consciously devoted--as a matter of clearly articulated policy--to educating the various publics about what various kinds of educational data mean, how they can be interpreted, and how variables among schools, districts, and states limit direct comparisons and often make them meaningless.

Offering wide range of indicators. The wider the range of indicators that is provided, both input and output indicators, the better the chance there will be of minimizing the misuse and misunderstanding of assessment and evaluation data. Obviously, there is no way within ordinary time and fiscal limitations that all input and output measures could be employed, but the broadening of the range is what is of importance. Insofar as the stated educational policy expands the range of the data made available, the chance is lessened that judgments will be made solely on the basis of some relatively simple indicator such as grade levels in reading, or SAT scores, or dropout rates.

Opting for full disclosure of data available. Hidden data is suspect data, and what is not known may be as dangerous as that which is misinterpreted. Therefore, there is every reason to believe that it makes better policy to display all the data that is available about the educational system, its failures and its successes, than to try to hold back that which might be "misinterpreted" or "misunderstood." Only the authorized policymakers can make the decision about what data to release, but every reasonable argument would seem to push that decision toward the fullest disclosure possible.

## IV. BUILDING COLLABORATIVE RELATIONSHIPS

The strong commitment of the SEAs, singly and collectively through the Council of Chief State School Officers, to seize the initiative in establishing a nationwide framework for educational evaluation and assessment commits them also to a high degree of collaborative effort. Federal, state, and local education agencies will all be deeply

involved. The specific procedural and operational details that will have to be worked out are not, however, the focus of this paper. Our concern here is only with policy issues, and only as these directly affect the SEAs, not the other partners who will be involved.

Recognition of different valid needs. At every level of the operation of the educational enterprise--federal, state, district, school-site, classroom, and individual student--there are varying needs for quite diverse kinds of data. There is also wide variance in the kinds of educational data which can legitimately be collected from each of these levels.

Development of a procedure for a reasonably complete and uniform data system recognizing valid needs and limitations at each level will require complex policy decisions and forthright policy statements, most of them probably limiting rather than expanding the scope of the data collection. Unless SEA policy is both clear and fair, suspicion and reluctance, rather than real collaboration, will likely dominate any attempt to develop useful comparative data.

Standardization and simplification. Sharing data among the various partners in the educational enterprise is going to require some perhaps painful modification of traditional beliefs and practices. Recognition of the uniqueness of each state, and a healthy respect for the virtues of local-board control of education sometimes make it difficult to see that somebody else's way of doing things may be OK, too! As data is shared, it will have to be standardized; and if it is to be used effectively, it will have to be simplified. SEA policy, it may be suggested, will need to reflect this willingness to standardize and simplify if real interstate collaboration is to be achieved.

Collaboration as action. It is not uncommon in any field--and education is no exception--to find calls for collaboration couched largely in rhetorical and hortatory terms. Collaboration does indeed require a psychological commitment--perhaps even an emotional one. It seems reasonable to suggest, however, that if educational evaluation and assessment is to become more of a nationwide effort, SEA policy needs to be stated not just in generally supportive terms, but in terms of specific action: commitments to be undertaken, changes to be made, goals to be achieved. This is the kind of policy which begets action.

Minimizing state-level burden on LEAs. Attempts to develop more comprehensive and more useful statewide evaluation and assessment systems--and through the state systems, at least an embryonic nationwide program--may founder on any number of rocky shoals, but nothing is more likely to threaten wreckage than the local suspicion that the state is burdening them excessively. Early in the game, as a matter of firm state policy, there needs to be hammered out common agreement on essentials: system outcomes expected; indicators to be chosen; methodologies to be employed; instruments to be used; funding patterns to be established; the sampling techniques to be followed wherever possible; and other procedural details. The details themselves, to be sure, are not policy as such; the policy comes in establishing the intended direction of minimizing state burden at the very outset of the program.

Perhaps the kind of policy suggested in the paragraph above has already been expressed in the CCSSO position paper, which calls for an evaluation and assessment system "as parsimonious and inexpensive as possible."

## IN CONCLUSION

As states prepare to increase the comprehensiveness of their statewide evaluation and assessment programs, and to structure these systems in such ways as to facilitate better and more fairly comparable nationwide reporting, a formidable number of technical questions remain unresolved. In addition, a host of new policy issues, forcing choices among policy options, are bound to emerge. Nevertheless, despite the uncertainties and even murkiness which surround the whole issue, one sure thing becomes apparent: the overriding importance of a flexible approach within a firmly-established policy framework. This will require a commitment to multiple purposes and multiple indicators but also an equal commitment to a sharply-focused emphasis on instructional improvement and program accountability.

1214s